Six-Code-Element Method of Numerically Encoding Chinese Characters And Its Keyboard

DESCRIPTION

[Para 1] In 1983, this inventor created WuBiZiXing technology, a universal system of encoding Chinese characters using the standard English keyboard, and obtained American, British and Chinese patents. That invention has solved the problem of efficiently inputting Chinese characters into computers, and become the dominant and most popular technology in this realm. But with the day-by-day growing demand for handling Chinese characters in other digital devices, such as mobile phones and PDAs, an easy and efficient method using numerical keys to input Chinese characters is universally desired.

[Para 2] This invention aims to solve the difficulties in learning and popularizing technology of encoding Chinese characters, and make it possible to encode Chinese characters with only numerical keys.

[Para 3] This invention relates to a universal system for encoding Chinese characters by using six code elements, and a kind of Chinese keyboard designed on the basis of the system. It can be realized entirely by using the six numeric keys on a numeric keypad of mobile phone, telephone or computer etc, to encode and input Chinese characters and Chinese words and phrases. The present invention is characterized in decomposing Chinese characters into six code elements: "一丨丿丶乙口" , which are respectively represented by six numbers "1 2 3 4 5 6" and in correspondence with the six numeric keys on a keyboard.

[Para 4] According to this invention, Chinese characters are regarded as a spelling of the above code elements. One can encode or keyboard a character in unit of code element in the order of handwriting. The code of a character

can comprise the character's whole code elements, or just include the first several and the last code elements. When a character happens to have elements less then the minimum number set in the system, the code comprises its whole code elements. For example:

[Para 5] The character: 啥: It can be decomposed into "口丿丶一一丨口". It's code for whole elements is 6341126, and the code for the method of encoding first four and the last code elements is 63416, and the code for the method of encoding first three and the last code elements is 6346. As for the character 中, which is decomposed into "口丨", for all the three encoding methods above, its code is 62.

In this invention, "一丨丿丶乙" are named as the five basic strokes. In each kind of strokes those similar in form are put together according to their writing order. Hereby "一" can also represents "ノ"; "丨" can also represent "丿"; "丶" can also represent "㇏"; "乙" can also represent all the various turning strokes as 乚乛𠃌 𠃍𠃋𠄎.

[Para 6] The existing technology is using 1, 2, 3, 4, 5 to represent "一丨丿丶乙". There are 5 strokes, and 5 numbers for encoding Chinese characters. On the basis of the existing technology, this present invention adds into a new code element "口" which corresponds to the numeric key 6, and becomes a new design. For example:

Character Examples	Codes based on The existing technology	Codes based on This invention
中	2512	62
和	3123251	312346
路	2512121354251	621213546
里	2511121	66121
电	25115	665

[Para 7] We can see from the above table, the lengths of the codes encoding by the six code elements of this invention are shorter than that of by the existing technology. To input these characters, the existing technology needs to strike 37 times of numeric keys, while this invention only needs 25.

[Para 8] This invention, only using five strokes and "口" for encoding Chinese characters and taking whole code elements or the first four and the last one as a code, is unprecedented in the realm of Chinese character encoding technology.

Through vast and numerous statistics and contrast researches on Chinese characters' components and their frequencies, the inventor discovered that the character-constituting frequency of "口" (including "日") is 34%, much higher than that of the other *compound components* (Chinese characters' geometrical elements containing two or more strokes, like 王，金，木，水，火，土). The total frequency of application of the Chinese characters which contain "口"(and "日") reaches as high as 44.35%.

[Para 9] Here is the statistical result of the appearance frequencies of the six code elements in 6763 Chinese characters (which constitute a character set as national standard GB2312-80):

Code Elements	Appearance Frequency **with 口**	Appearance Frequency **without 口**
一	18,459	21,870
丨	11,061	13,728
丿	11,495	11,495
丶	12,012	12,012
乙	10,054	12,721
口	3,411	0

[Para 10] This is not only the reason that this invention chooses only "口" but not other components as the code element, but also the essential reason that this invention has a substantial advantage of practicability comparing with the existing technology. This invention cannot be deduced simply from the existing technology. Data in the comparative table below is the important basis for optimally selecting code elements and cannot be predicted by anybody without creative work.

Comparative Table of Total Frequency of Components in the Most Commonly Used 1000 Chinese Characters

Order	Components	Character-constituting Frequency%	Application Frequency%
1	**口**	**34.00**	**44.35**
2	人	7.70	9.36
3	土	8.70	7.74
4	勹	1.10	5.31
5	亻	5.70	5.13
6	大	4.00	4.92
7	二	4.60	4.40
8	又	4.80	4.04
9	冂	4.60	4.01
10	厶	3.40	3.83
11	木	5.90	3.64
12	小	3.40	3.61
13	十	4.50	3.55
14	月	3.30	3.34
15	辶	2.60	2.92
16	扌	4.10	2.85
17	八	4.20	2.77
18	寸	2.20	2.62
19	犭	1.20	2.55
20	匕	2.20	2.55
21	氵	1.20	2.48
22	艹	4.00	2.48

[Para 11] The above research result shows that, "口" has the highest character-constituting and application frequencies among all the compound components of Chinese character. Therefore, optimally selecting "口" as a new code element will effectively shorten the length of codes, reduce key-press

times, and considerably increase the uniqueness of code and input efficiency. This is a creative design of this invention. The meaning of "口" in this invention is just as important as the nib to a pen.

In addition, according to this invention, When encoding the most commonly used Chinese characters like "的、是、和、中、国" , 口 (and "日") don't need to be decomposed into single strokes. As a result, not only the process of inputting the most commonly used Chinese characters is considerably simplified, but also the identical codes are greatly reduced, as shown in the table below (Identical codes are for the first six digits):

Chinese characters	The existing technology		This invention		
	Codes	Other Characters with identical codes	Encoding whole elements	Encoding the first four and the last one	Other Characters with identical codes
的	32511354	卑伊鬼魄兜皈皑皓魁龄……	366354	36634	None
和	31234251	利秒种租积犁程稍香秈……	312346	31236	种积程稆稞香馥
中	2512	串典吊同吕吵员咒品哭……	620	620	None
是	251112134	匙呈味吁早时果晒晶罡……	6612134	66124	匙晷题题
国	25112141	哇吐啪啦啧里野哳嗑嗜……	611214	61124	逞

[Para 12] It can be seen from the examples above that the existing technology has too many identical codes, while there are no or very few identical codes when using this invention to encode these characters.

[Para 13] When we encode 6763 characters in China's national standard character set GB2312-80, comparative table of "Code uniqueness" between this invention and the existing technology can be shown as:

	Characters with no identical codes		Characters with no identical codes +Characters with 2 identical codes +Characters with 3 identical codes	
	Characters	Proportion	Characters	Proportion
The existing technology	428	6.33%	428+392+294=1114	16.47%
This invention	730	10.79%	730+602+444=1776	26.26%
Conclusion	The code uniqueness of this invention is 70% higher than that of the existing technology.		The code uniqueness of this invention is 59% higher than that of the existing technology.	

[Para 14] It can be seen that this invention has an obvious advantage in terms of practicability because of its code uniqueness. Compared with the existing technology, this invention has made an important technical progress.

[Para 15] In addition, there are 96 characters which contain "口" and "日" in the 500 commonly used characters, and they hold 19% of these 500. Because these characters have the highest frequency of application, this invention improves their code uniqueness, thus definitely has more outstanding practicability than the existing technology.

[Para 16] Compared with the existing technology, this invention sacrifices very little in terms of easy to learn, because it has only added into one more code element and used one more key. But the substantial technical progress,

which is made by this invention, is very obvious. This is the creativeness and practical value of this invention.

[Para 17] This invention also characterizes in that when using the six code elements "一丨丿丶乙口" to input simplified/traditional Chinese characters in the order of handwriting, the encoding can be completed either when the character just appears on the screen, or when the character's whole code elements are inputted.

[Para 18] In order to abridge the codes, this invention allows to select part of a character's code elements, that is, only select the character's first several, and the last several or one code elements for encoding. For example, selecting a character's first 5 and the last 1 code elements, or selecting its first 4 and the last 1 code elements, or selecting its first 3 and the last 1 code elements, or selecting its first 4 and the last 2 code elements, or selecting its first 3 and the last 2 code elements to encode and input the Chinese character by numerical keys.

Chinese characters forms can be classified by the information of their forms into two basic topological patterns, namely, Compound and Singular. Compound topological-patterned character can be divided into at least two parts visually, like 洋，音，边. While single topological-patterned character can't be divided, such as 中，乐，土. According to this invention, when encoding the characters, as for the compound, one can divide it into two parts, and just encode the first and the last code elements of its first part, and then encode the first three and the last code elements of the second part, so the maximum length of a compound character's code is six. As for the single topological-patterned character, one just needs to encode its first four and the last code elements, and the maximum length of code is five.

[Para 19] According to this invention, the most commonly used character component "口" is encoded as "6". Based on this, the component "日" can be regarded as two "口". So "日" can be encoded as "66". For example, the code of "旦" is 661; the code of "时" is 66124; and the code of "电" is 665.

[Para 20] In this invention, considering character component's derivation and its intuitional meanings, the component "口" in the character "国" is also encoded as 6. Thus, for example, "国" is encoded as 611214; "回" is encoded as 66; "因" is encoded as 6134.

[Para 21] In the process of the key-in of a character, in case of identical codes, all the characters are ordered by the frequency of application. A more frequently used character will first appear at the right position of the line on the screen.

This invention can be used to handle both simplified/traditional characters and words and phrases. When inputting phrases, one can switch (for example, press "*" key to signal) the system into a state of only-phrase inputting, or ignore the states to mix the single character and words and phrases to input.

[Para 22] There are various and flexible ways of encoding phrases, such as selecting 2-4 code elements from each character of a 2-character phrase, selecting 2-3 code elements from each character of a 3-character phrase, selecting 2 code elements from each character of a 4-or-more-character phrase, or, selecting 2-3 code elements from the first two and the last characters of a 3-or-more-character phrase. For example:

2-character phrase:

经济——554414 (method 1: 经: first 2 elements + 济 first 4 elements)

经济——551441 (method 2: 经: first 3 elements + 济 first 3 elements)

3-character phrase:

Simplified: 电视台——664554 (first 2 elements for 电,视,台 respectively)

Traditional: 電視臺——144512 (first 2 elements for 電,視,臺, respectively)

Multiple-character phrase:

中华人民共和国——623261 (first 2 elements for 中,华,国 respectively)

香港特别行政区——*314413 (first 2 elements for 香,港,区 respectively)

[Para 23] Since the method of encoding phrases is choosing the first several code elements (most of them are roots of Chinese characters) of each character, so the codes in this invention have been well dispersed and can avoid identical codes between phrases and single characters. For example, selecting the first three code elements from each character of "河流" thus its code is "441441". Because there is no character which contains two "氵" (a root of Chinese character), this phrase will not have identical code with single characters. This design makes it possible to input single characters and phrases together. It is a creativeness of this invention.

[Para 24] This invention also characterizes in its simple and easy-to-remember rules. Generally, one who can write Chinese characters is able to master this method within ten minutes.

[Para 25] The distribution of the numeric keys used in this invention can be in the way of a telephone keypad, namely, "1, 2, 3" are distributed on the top row of the keypad; and the numeric keys also can be distributed according to the PC numeric keyboard, namely, "1, 2, 3" are on the bottom row. And no matter adopting what kind of key distribution, the five basic strokes and "□" can be printed or carved on the six numeric keys 1, 2, 3, 4, 5, 6.

[Para 26] This invention can be used to encode and input all simplified/traditional Chinese characters in any character sets.

[Para 27] This invention is also a creative method of sorting and searching Chinese characters in dictionaries. The process is: encode all the Chinese characters and phrases into numbers by this invention, and then sort the Chinese characters in the increasing order of their codes, and make it be an index of Chinese characters and words and phrases in a dictionary. This is going to be a more practical, easier and quicker character-searching method than any of the existing ones.

[Para 28] The method of encoding Chinese characters by this invention can be brought into the primary or middle school education over the countries and areas where using Chinese characters. It can be designed into many kinds of teaching materials and software in order to let children know each character's correct writing order and know how to input them into computer, mobile phone and other digital devices.

[Para 29] After encoding all Chinese characters and words and phrases according to this invention, we can design the input software for computers and mobile phones, and character-searching software depending on input data. Thereafter this invention can be applied onto all kinds of communication and special products that need to input Chinese characters with numeric keypads, such as mobile phone, computer, and Chinese PDA, etc.

[Para 30] The great progress made by this invention can be illuminated in Table 1. This table shows the comparative results between various existing mobile-phone-Chinese-character-input methods with this invention. When we use all these methods to input 1000 most commonly used Chinese characters, it can be found that this invention needs the least average key-press times. So obviously this invention is the most efficient technology.

[Para 31] The design of this invention's keyboard is shown in Figure 1. Case A is how the numeric keys distribute on PC keyboard, and Case B is how they distribute on mobile phone and telephone' keypads. Different distributions do not affect on the substantive characteristics of this invention.

[Para 32] When this invention is realized on PC, the brief flow chart of the Chinese-character-searching software is shown in Figure 2.

Table 1: Comparison of Key-Press Times Among Various Methods

(Encoding 1000 Most Commonly Used Chinese Characters)

(Times of Pressing Keys)

No.	CHARAC.	This Invention		Existing Mobile-Phone-Chinese-Character Input Method				
		Whole Elements	First four & Last one	Nokia (5 keys)	Motorola (iTAP)	KingJia (9 keys)	Hai'er (8 keys)	Samsung (19)
		Average 4.6	Average 4.3	Average 6.7	Average 6.1	Average 6.6	Average 6.3	Average 5.1
1	的	1	1	2	2	3	3	1
35	成	3	3	6	6	5	4	3
70	去	3	3	5	3	3	6	3
105	里	4	4	5	5	6	5	4
140	用	5	5	7	5	6	6	5
175	发	4	4	6	5	3	5	4
210	见	4	4	5	5	6	5	4
245	回	4	4	5	4	5	5	4
280	全	4	4	7	7	7	7	5
315	分	5	5	5	5	6	5	5
350	少	4	4	6	6	3	6	4
385	市	5	5	6	6	6	6	6
420	市	4	4	6	4	6	6	4
455	住	5	5	7	6	7	7	4
490	效	6	6	8	7	8	7	6
525	那	5	5	6	7	8	6	5
560	两	6	5	8	8	7	8	6
595	如	5	4	8	6	8	8	6
630	时	5	4	7	7	7	7	6
665	吐	4	4	7	7	6	6	6
700	代	4	4	7	7	8	6	5
735	场	5	5	9	8	10	8	7
770	公	5	5	7	5	6	7	6
805	报	5	5	9	8	9	9	7
840	发	5	5	11	8	11	11	4
875	教	4	4	8	7	9	7	7
910	老	5	5	9	10	8	9	8
945	现	7	5	11	11	9	11	9
980	保	6	5	7	6	7	7	6
1000	来	5	5	7	7	7	7	5